User Guide for USDA-C-Grain-sum software

Written by: Nicholas Sixbury

USDA-ARS Manhattan, KS

May / 2024

The C-Grain imaging system is used to rapidly take images from cereal seed samples and output size and color parameter information per seed or per sample. It can output several types of data files, such as CSV format or XML format. In the CSV output files, information is listed for each kernel, with a number of measurements provided. In the XML output files, measurements are only listed per sample (composed of many kernels), and some machine learning results from each sample are also included.

This software was written at USDA-ARS of Manhattan by Nicholas Sixbury in the spring and summer of 2024, developed for use by scientists Scott Bean and Rhett Kaufman. The purpose of the software is to extract the useful information from the CSV and XML files created by the C-Grain in order to save time that would otherwise be spent manually processing the files in excel. The goal of the project is to create a summary of each sample, with each sample containing a number of kernels. This summary is created as an excel file with each sheet holding different information on each sample.

How to use the summary software:

1. **Open the program** from the desktop by using the mouse to click on the icon.

Figure 01: example of shortcut to software.

2. The configuration preset will be set to Wheat, Sorghum, or Other. The configuration sets the desired output, which is sent to an excel xlsx file.
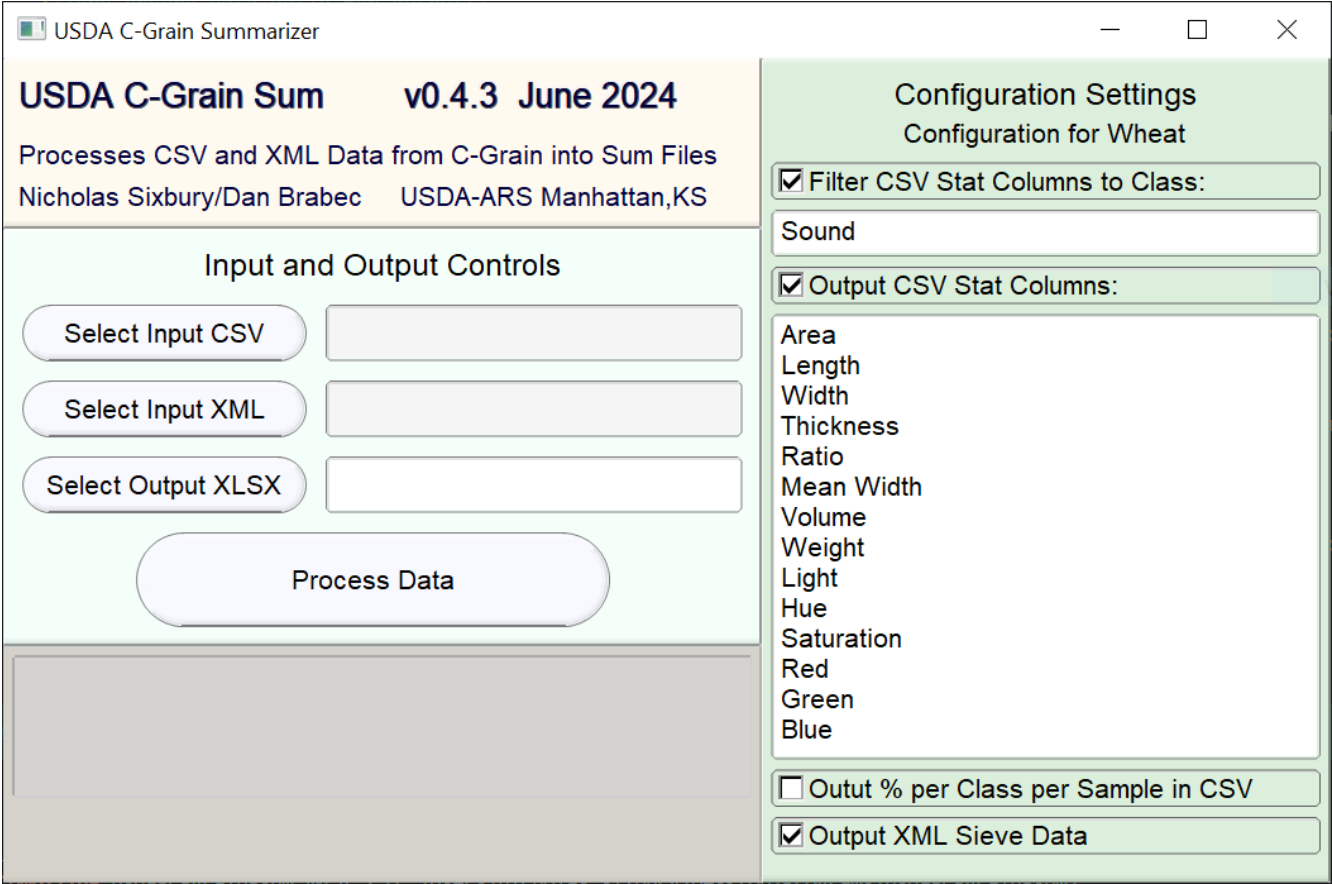

Figure 02: example of software GUI with configurations options on the right side.

3. Choose input files by clicking "**Select Input CSV**" or "**Select Input XML**". If the Sorghum Config is selected, then you only need to select a CSV file, whereas if the Wheat Config is selected, then you must select both a CSV and an XML file.

4. Choose output file name by clicking "**Select Output XLSX**" or just typing a name in the box. The program will automatically set the file extension.

5. Click "Process Data" to create the output file from the inputs you selected. After the processing has finished, a dialog will appear, asking if you'd like to open the folder where the output file is located. Selecting "yes" in this dialog will show you directly to the output file you generated.
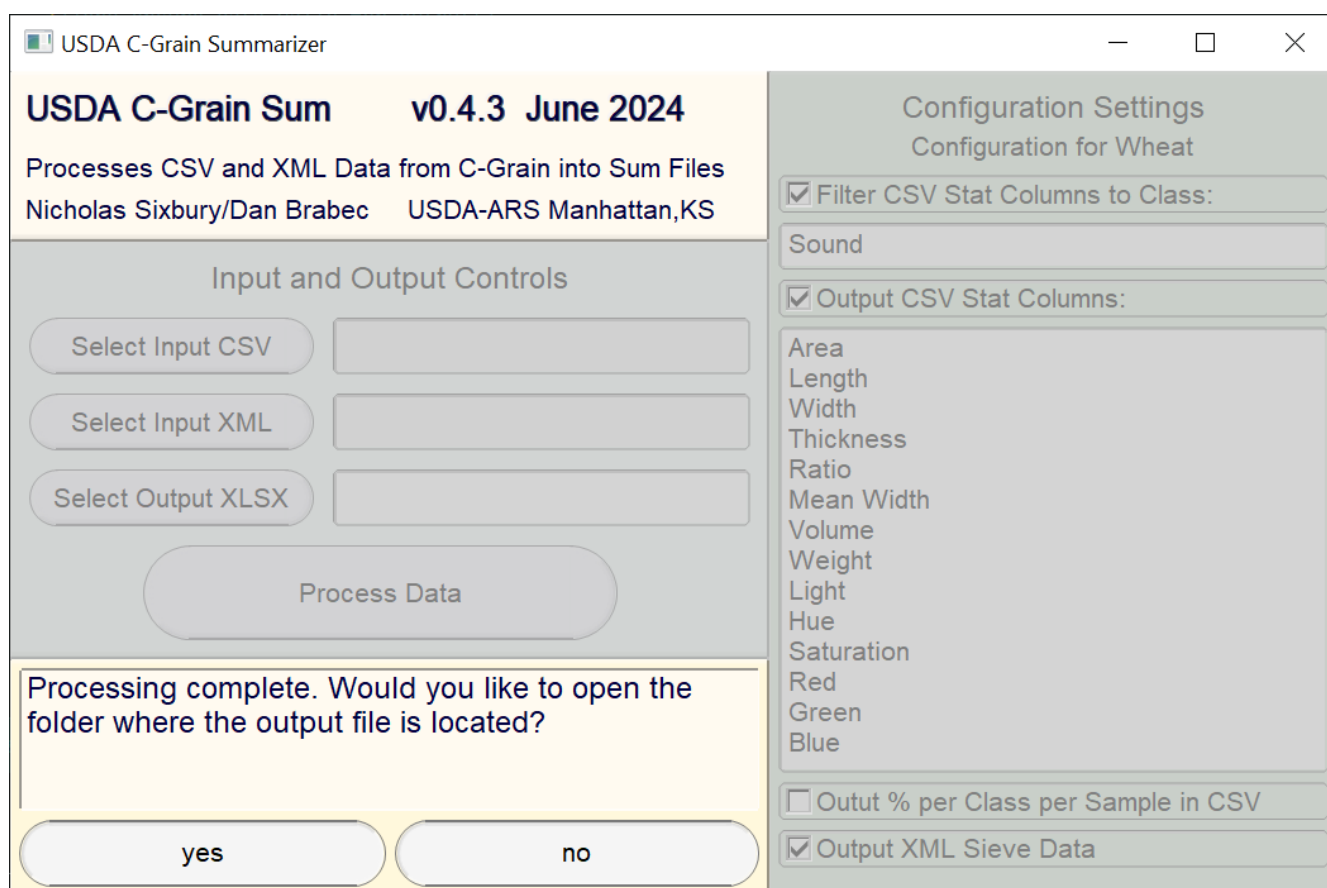

Figure 03: example of software GUI, showing dialog to show output folder in bottom left

6. Once the program has finished processing the file, the section labeled "Input and Output Controls" will be cleared, and you can continue processing as many files as you like.

Advanced Options for Sum Software:

Reset the Configuration Settings

If, at any time, you wish to reset the configuration settings, simply right-click the label "Configuration Options" to the right. A dialog box will pop up asking if you're sure you want to continue. Click yes, and another dialog box will appear, asking which preset you'd like. Select any option, and your configuration will be reset to the option you choose.

Change the Classification Filter for CSV Stat Columns

In order to enable class filtering for csv stat columns, the checkbox labeled "Filter CSV Stat Columns to Class" must be checked. By the default, the box to the right will say either "Sound" or "Sorghum". The way this classification works is that it removes kernels with classes not specified from consideration. So, if "Sound" is specified, then csv stat columns will only consider kernels with a classification of "Sound". If you wish to filter for multiple classes, such as "Sound", or "Black-Germ", then you can enter however many class names you wish, separated by the | character. If multiple classes are specified, then kernels with any of the classes specified will be considered. By default, in order to determine what a kernel is classified as, the program looks for the column in the csv labeled "raw-filtered-as". If this column cannot be found, then it will default to looking in the 6th column.

Change the Column Config for CSV Stat Columns

The csv stat columns output option calculates the average and standard deviation for several columns of data in the csv. The way this is configured is quite simple: If the text "Area" is placed in the column config, then the program will look for a column called "Area" in the csv. Each value in the config can be placed on a different line. If you wish to see a list of common columns names, simply click the config area, which lists all the column names, and press the F1 key.

Changing the Sample ID Grouping in the Output for CSV Files

By default, in order to determine the sample id of a particular kernel, the program looks for the column labeled "external-sample-id", and if that column cannot be found, then the program defaults to using the 3rd column, which should have the label "external-sample-id" by default. If you wish to use a different column for grouping samples, simply right click on the button labeled "Select Input CSV". This will open a dialog allowing you to change which column the program will look for to determine sample id. If the column header you select can't be found, then the program will attempt to fall back to a default.

Configuring Advanced XML Options

There are a few advanced options relating to how the program parses XML files, available by right clicking the button labeled "Select Input XML". By default, the program looks for a tag labeled "reference" to find what in the CSV is called "external-sample-id". If you wish for the program to use a different tag as the sample id (which determines how results will be grouped), then you can change this using the box provided. Additionally, in the XML files, a tag called "sample-result" wraps each sample, so the program looks for this tag and knows that it's reached the end of a sample if it finds a closing tag for "sample-result". If your XML file is a little different, and samples are wrapped in a different tag, you can edit this value in order to make the program work with your input. Additionally, by default, the program only reads the sample-id and tags starting with "filter-sieving" from the XML, ignoring any tags which don't have data in them. If you wish to include additional tags in the output, you can type the names of all the tags you want to include in the box at the right.

Troubleshooting:

When opened, Program displays a dialog message saying something about config:

The program saves certain configuration information to a file in order to save your settings when the program is closed. The file is located in the same directory as the executable file for portability. If the config file cannot be found, the program assumes that you opening the program for the first time, so it will ask you if you want to apply a preset. If you don't want to bother with it, simply click "no", and a default config will be created. Once this process completes, and you close the program once, you shouldn't see the message again. If you continue seeing the message every time you open the program, this indicates that the application is encountered difficulties when saving the file. If this occurs, you can still use the program normally, but please contact the developer for support in order to resolve the issue.

After clicking "Process Output", Program displays series of Error dialogs and closes:

The most likely cause of this problem is an invalid output path name. Each operating system has certain characters that cannot be included in a filename. If you type your output filename into the box manually, then there is no check preventing you from creating an invalid output path, thus when the program tries to write to the path, it encounters an error. This can be fixed by not including illegal characters, which is made easier when you use the file dialog opened by "Select Output XLSX" to select your output file, as you will be warned when trying to create an invalid filename.